Scheduling Enhancements and Performance Evaluation of Downlink 5G Time-Sensitive Communications

RENATO B. ABREU[1], **(Member, IEEE), GUILLERMO POCOVI**[1], **(Member, IEEE),**
THOMAS H. JACOBSEN[1], **(Member, IEEE), MARCO CENTENARO**[2], **(Member, IEEE),**
KLAUS I. PEDERSEN[1,2], **(Senior Member, IEEE), AND TROELS E. KOLDING**[1], **(Member, IEEE)**

[1]Nokia Bell Labs, 9220 Aalborg East, Denmark
[2]Department of Electronic Systems, Aalborg University, 9220 Aalborg East, Denmark

Corresponding author: Renato B. Abreu (renato.abreu@nokia-bell-labs.com)

ABSTRACT Time-sensitive communications (TSC) in wireless networks is an emerging paradigm that gains research momentum as an enabler of the industrial Internet of Things. As compared to ultra-reliable low-latency communications (URLLC) in fifth-generation cellular networks, TSC has stricter requirements in terms of latency and reliability, and also demands absolute time-synchronization and on-time delivery of packets for deterministic and isochronous real-time applications. In this regard, a key question is how to schedule TSC traffic flows effectively. This paper presents a radio resource allocation strategy for deterministic downlink TSC flows leveraging traffic pattern knowledge. Taking physical layer control channel effects into account, a comparison of semi-persistent and dynamic packet scheduling methods is presented as well as necessary enhancements to link adaptation and interference coordination procedures. With the proposed methods, the network capacity in terms of number of supported TSC flows can be more than doubled compared to traditional dynamic scheduling methods as commonly assumed for URLLC applications.

INDEX TERMS 5G, time-sensitive communications, time-sensitive networks, Industry 4.0, packet scheduling, semi-persistent scheduling, dynamic scheduling, link adaptation.

I. INTRODUCTION

The fourth industrial revolution, also known as "Industry 4.0", sets out to significantly improve productivity and efficiency of manufacturing and industrial processes. Use cases of Industry 4.0 include isochronous motion control systems, advanced audio/video production, and synchronized collaboration in large swarm systems, including seamless cooperation among intelligent robots and humans [1]. The communication layer is an essential part of this vision, delivering time sensitive data among all collaborating and networked components. State-of-the-art wired industrial Ethernet can provide time synchronization and tight gating mechanisms to allow strict and time-bounded message delivery of time-sensitive traffic flows e.g. sensor data, control input to actuators, and audio or video packets [2]. A key

industrial Ethernet standard, Time-Sensitive Networking (TSN), is standardized in IEEE as a part of the 802.1Q standard family, see e.g., [3] and embedded references.

The next tier of efficiency enhancements in Industry 4.0 is envisioned to come from wireless communications, enabling more flexible reconfiguration of the production environment and reducing cabling costs [4]. The available industrial wireless communication solutions, e.g. ISA100.11, WirelessHART, and WISA [5], are mainly used in the context of wireless sensor networks or human machine interfaces, and do not scale to the requirements of TSN networking. To improve the scalability as well as to mitigate the reliability limitation of the existing technologies by leveraging the benefits of licensed wireless spectrum, the third generation partnership project (3GPP) specification group has, with its fifth generation (5G) New Radio (NR) wireless standard, significantly increased the commitment on the topic, here referred to as industrial Internet of Things (IIoT).

The associate editor coordinating the review of this manuscript and approving it for publication was Giacomo Verticale.

In 5G NR Release-15, the support for ultra-reliable low-latency communications (URLLC) was introduced, guaranteeing a one-way message delivery delay of less than one millisecond with five nines of reliability (99.999%). Among others, this is achieved thanks to flexible frame structure supporting short transmission-time intervals (TTI) and reduced base station and terminal processing times to meet the low-latency requirement, as well as methods to achieve improved reliability such as packet duplication, lower modulation and coding schemes (MCSs) and logical channel priority restrictions [6], [7]. The support for end-to-end (E2E) deterministic communication required by TSN services was provided by 5G NR Release-16 with the introduction of the so-called time-sensitive communication (TSC) [8], [9]. While URLLC traffic is characterized by a random packet generation process, often experiencing time intervals of several milliseconds between two adjacent packet arrivals, TSC is characterized by deterministic traffic with fixed inter-packet arrival times such as 0.5 ms, and packets must be delivered according to an agreed time-schedule with microsecond resolution. To meet TSC requirements, the radio access network must support a one-way latency down to 0.5 ms with up to six-nines reliability in terms of packet error rate. As an example of TSC service requirements, Table 1 provides the specific values for motion control use cases in terms of service availability, packet periodicity, maximum packet latency, payload size, and number of supported UEs [8].

TABLE 1. Requirements for motion control use cases of IIoT [8]. The number of UEs is normalized for an area of 50 x 10 x 10 m.

Case	Availability	Periodicity	Latency	Payload	UEs
I	99.999% to 99.99999%	0.5 ms	0.5 ms	50 bytes	20
II	99.9999% to 99.999999%	1 ms	1 ms	40 bytes	50
III	99.9999% to 99.999999%	2 ms	2 ms	20 bytes	100

One key enabler of TSC in 5G systems is the introduction of TSC Assistance Information (TSCAI) in the radio access network (RAN), which provides information on absolute time offset, the periodicity, and the direction (uplink or downlink) of the traffic arrivals of each TSC traffic flow. This information can be used to adjust the resource allocation in the 5G base station (gNB) in line with the traffic characteristics. This is especially the case for configured grant (CG) scheduling in uplink and semi-persistent scheduling (SPS) in downlink, but also valid for TSCAI-aware dynamic packet scheduling (DPS) schemes. In particular, SPS and CG based schemes reduce the dependency on the downlink control channel as compared to DPS procedures [10], and are attractive for reaching high reliability and extremely low delay. SPS and CG operation have been significantly enhanced in Release-16; for instance, a UE can be configured with multiple independent SPS/CG allocations tailored to

different TSC flows, and the minimum SPS periodicity has been reduced from 10 radio slots in Release-15 to one radio slot (e.g. 0.5 ms for 30 kHz sub-carrier spacing, SCS). We refer to [11], [12] for a comprehensive overview of the IIoT-related enhancements in Release-16.

A. SCHEDULING CHALLENGES AND RELATED PRIOR-ART
Despite these noteworthy enhancements, there are still various unsolved challenges related to efficient link adaptation and scheduling of TSC traffic in the RAN. For instance, due to the strict TSC latency requirement of 0.5 ms, HARQ retransmissions cannot be used as a method to improve reliability. In other words, the targeted 99.999% reliability needs to be achieved with a single over-the-air transmission, which calls for extremely accurate link adaptation. This problem has been studied for the case of DPS procedures for URLLC traffic [13], [14], [15]. In [16], a resource reservation scheme is proposed to reduce delays caused by scheduling request for bursty traffic. A fixed MCS is assumed for all UEs in this study. In [17], resource allocation is optimized for reducing the packet loss probability for bursty URLLC traffic, while scheduling with neighbor cells interference is not considered. The problem however is still open for SPS where slow link adaptation is applied, i.e. the MCS is not adjusted for every transmission (as for DPS), but is instead selected for a longer time period per user. The selected MCS should be as high as possible to reduce radio resource consumption and correspondingly increase the number of supported users in the system, subject to fulfilling the strict outage requirements.

Semi-persistent resource allocation techniques has been mainly studied for transmission of Voice over Internet Protocol (VoIP) traffic. For instance, the work in [18], [19] demonstrate the higher performance of SPS for VoIP compared to DPS. However, due to the TSC traffic considered in our study with much stricter system requirements and less dynamic behaviour, the findings of VoIP-related studies [18], [19] are not easily applied to our case. More recently, some studies analyze the applicability of SPS-alike techniques for scheduling machine-type communications (MTC), where the objective was to minimize the number of frequency bands used by the MTC devices to allow flexible resource allocation for other types of traffic [20]. Other studies, such as [21], [22], tackle the problem of *massive access* and generally assume more relaxed quality of service (QoS) requirements. For deterministic IIoT traffic, a simplified single-user analysis has been conducted by different companies in 3GPP (summarized in [9]), bringing to light the feasible physical-layer configurations (SCS and TTI duration) and MCSs to achieve the target latency and reliability for certain signal quality conditions. To the best of our knowledge, there are no existing studies addressing the problem of radio resource allocation for periodic industrial traffic with very short periodicity and extreme requirements of 0.5 ms and five- to six-nines reliability in a complex dynamic multi-user and multi-cell wireless network.

B. CONTRIBUTIONS OF THIS PAPER

The core of this paper consists of a thorough and systematic evaluation of the performance of different scheduling strategies in the novel context of 5G-enabled TSN, that is, TSC. In particular, we present a resource allocation strategy of downlink TSC traffic flows, leveraging the information provided by the TSCAI to optimize the radio resource utilization to maximize the number of supported TSC users in the network. Solutions are presented for both the case of DPS and the case of SPS in a dense indoor cellular network. The contribution of this paper is threefold:

1) We present an enhanced link adaptation mechanism tailored to achieve the 0.5 ms latency requirement with 99.999% reliability with a single transmission attempt. For the case of DPS, we build on existing URLLC link adaptation studies [13], [14]; whereas, a significantly different novel solution is presented for SPS.

2) Interference-aware technique for SPS resource block (RB) allocation is presented, targeting to minimize the experienced inter-cell interference, which is a major problem in the considered dense indoor scenario.

3) Finally, by taking into account control channel non-idealities, a detailed comparison of DPS and SPS methods is presented, highlighting the benefits of the proposed solutions. It is shown that the network capacity in terms of number of supported TSC flows can be more than doubled compared to scheduling approaches typically assumed for URLLC applications.

The performance of the proposed solutions satisfying TSC constraints is assessed by means of extensive dynamic system-level simulations aligned with 3GPP NR evaluation assumptions. The simulations are designed to be highly realistic, with accurate modelling of the main performance-determining radio protocol effects of dynamic multi-user multi-cell networks, which would be hard to capture with analytical models (such as in [17] which does not account for interference from neighbor cells). To the best of our knowledge, this study is the first of its kind as far as TSC is concerned, as opposed to studies on URLLC.

The rest of the paper is organized as follows. In Section II the network model details are introduced, as well as the problem formulation. Section III presents the studied schemes with the envisioned enhancements for improving the 5G network capacity when serving multiple users with TSC traffic. Section V present details of the evaluation methodology and the dynamic system-level simulations assumptions. Finally, the results of this study are provided in Section VI, followed by the conclusions in Section VII.

II. SETTING THE SCENE
A. NETWORK AND TRAFFIC MODEL

We assume a multi-user multi-cell synchronous network comprising a set of cells $\mathcal{C}$, of cardinality $|\mathcal{C}| = C$, as illustrated in Fig. 1 for $C = 12$. Note that we use the terms *cell* and *gNB* interchangeably throughout the paper,
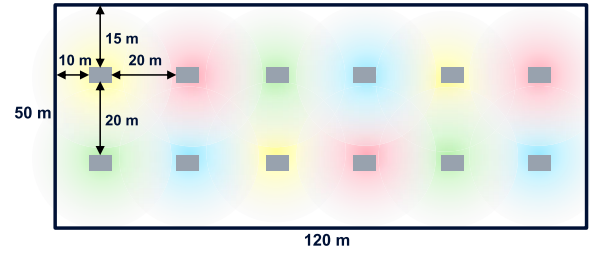

FIGURE 1. Layout of indoor deployment with gNBs positions, represented by the grey rectangles.

since sectorization is not employed. A set of semi-stationary UEs $\mathcal{U}$, of cardinality $|\mathcal{U}| = U$, are randomly distributed in the network according to a spatial uniform distribution. The serving cell for each UE is the one with the highest receive power, and the UE remains connected and time-frequency synchronized with the serving cell throughout the entire call.

A downlink TSC traffic flow is generated for each UE with a fixed periodicity T, latency requirement D, and payload size B. We assume the most stringent requirements from Table 1 with a periodicity of $T = 0.5$ ms and a latency of $D = 0.5$ ms, while different values of B are considered for evaluation. The core network provides information to the RAN on the QoS requirements that shall be fulfilled. Specifically, each TSC flow is assigned with a 5G QoS Indicator (5QI) for delay-critical guaranteed bit rate (GBR), where the payload size B is given by the maximum data burst volume (MDBV) parameter, and the latency requirement D is provided by the packet delay budget (PDB) parameter [10]. On top of this, TSCAI provides information on absolute time offset of payload arrival and periodicity T. The gNBs use this information for making resource management decisions such as aligning the resource allocation to the traffic arrivals.

B. RADIO FRAME STRUCTURE & RESOURCE ALLOCATION

Users are scheduled on a time-frequency grid of resources, using orthogonal frequency division multiple access (OFDMA) and frequency-division duplexing (FDD). A physical-layer numerology configuration with 30 kHz SCS is assumed, where the scheduling resolution consists of an RB of 12 sub-carriers in the frequency-domain and a mini-slot of 2 OFDM symbols (i.e. TTI duration of $t_{TTI} = 71$ μs) in time-domain, resulting in 24 resource elements (RE) per RB. A RE is the smallest unit of the resource grid, i.e., one OFDM symbol in the time domain and one sub-carrier in the frequency domain. The bandwidth contains Y RBs in total.

Each data transmission to a user can be either dynamically allocated with DPS or semi-statically allocated using SPS by the gNB in the time-frequency grid. With DPS, the gNB sends a downlink control information (DCI) to the UE on the physical downlink control channel (PDCCH), determining the time-frequency resources of the data allocation, the employed MCS, as well as other DL transmission parameters to be used. The gNB selects the aggregation level (i.e. effective coding rate) of the PDCCH in accordance with

the user's channel condition and the target control channel reliability. Specifically, the DCI is transmitted on a number of control-channel elements (CCE) in the range {1, 2, 4, 8, 16}, where each CCE consists of 6 RBs in frequency and 1 OFDM symbol in time, i.e. corresponding to 72 REs.

For the semi-static allocation scheme, the gNB provides the UE with an SPS transmission periodicity via Radio Resource Control (RRC) signaling [23]. The SPS transmission is activated by transmitting a DCI with special field settings [24] that express the frequency resources, MCS, and other DL transmission parameters in a similar way as for dynamic scheduling. Upon activation of the SPS, the UE receives DL data with the periodicity provided by RRC signaling and transmission parameters according to the activation DCI. A new DCI can be transmitted to adjust the SPS transmission parameters, as well as to release the semi-persistent allocation.

C. LATENCY BUDGET ANALYSIS
The latency of each TSC payload in the RAN is measured from the moment it arrives at the serving gNB until it is successfully received at the UE. Assuming that the TSC payload can be entirely scheduled on a single TTI, the latency of a successfully-received first transmission τ_1 equals

$$\tau_1 = t_{\text{queue}} + t_{\text{gNB}} + t_{\text{align}} + t_{\text{TTI}} + t_{\text{UE}} , \qquad (1)$$

where t_{queue} is the queuing delay of the TSC payload at the gNB; $t_{\text{align}} \in [0, t_{\text{TTI}}]$ is the so-called TTI alignment; and t_{gNB} and t_{UE} is the processing time at the gNB and UE, respectively. Fast UE and gNB processing capabilities are considered, corresponding to $t_{\text{gNB}} + t_{\text{UE}} = 214 \, \mu s$ (6 OFDM symbols) [25], [26]. For networks with low offered load, $t_{\text{queue}} \to 0$, thus the experienced latency is equal to

$$\tau_1 = t_{\text{gNB}} + t_{\text{align}} + t_{\text{TTI}} + t_{\text{UE}} = 357 \mu s . \qquad (2)$$

This means that the TSC latency requirement $D = 0.5$ ms can be fulfilled conditioned on the correct reception of a single transmission attempt, preventing further HARQ retransmissions as they would increase the latency to approximately $893 \, \mu s$ [25]. On the other hand, we note that the latency budget τ_1 gives still some room to handle potential queuing delays $t_{\text{queue}} > 0$ in realistic situations (e.g., in presence of a high TSC offered load an low instantaneous SINR for a UE); these aspects are taken into account in the performance evaluation.

III. LINK ADAPTATION ENHANCEMENTS FOR DYNAMIC PACKET SCHEDULING
In this section we present enhanced link adaptation strategies for DPS targeting delivery of TSC payloads with a single transmission opportunity.

A. BASIC LINK ADAPTATION
The downlink dynamic packet scheduler exploits the channel quality indicator (CQI) reports from the UE for performing

resource allocation and MCS selection for individual data transmission [26], [27], [28]. The CQI is periodically reported by the UE every N TTIs with a sub-band resolution of $y \leq Y$ RBs in frequency. For each sub-band s, the reported CQI indicates the highest supported MCS index m_s^* which satisfies the condition

$$m_s^* = \arg \max_m \{R_m | P_e(\Psi_s) \leq P_{\text{target}}\} , \qquad (3)$$

corresponding to the m-th MCS index [26, Sec. 5.2.2.1] that provides the largest data rate, R_m, with a block error probability (BLEP) P_e not exceeding P_{target}. In practice, the UE i) measures the experienced Signal to Interference and Noise Ratio (SINR) Ψ_s on each sub-band s; ii) estimates P_e for each of the supported MCSs, given its knowledge of the BLEP vs SINR mapping curves, and iii) selects the CQI by comparing P_e with the threshold P_{target}, which in NR can be either 10^{-1} or 10^{-5} [26]. In particular, $P_{\text{target}} = 10^{-5}$ is adopted in this work due to the stringent reliability requirement of TSC.

The post-receiver SINR is used for deriving the CQI report, as it considers the precoding and combining gain of multiple N_t and N_r transmit and receive antennas, respectively. The post-receiver SINR measurement on the n-th TTI and s-th sub-band is given by,

$$\Psi_s[n] = \frac{\Omega_c[n] \|\mathbf{g}_s[n]\mathbf{H}_{c,s}[n]\mathbf{f}_c[n]\|^2 P_c}{\sum_{i \in \mathcal{I}} \Omega_i[n] \|\mathbf{g}_s[n]\mathbf{H}_{i,s}[n]\mathbf{f}_i[n]\|^2 P_i[n] + \sigma_0^2} , \qquad (4)$$

where sub-index c denotes the serving cell; $\mathcal{I} \subseteq \mathcal{C}$ is the set of cells that create interference to the UE; Ω_i denotes the large scale fading (path-loss and shadowing); $\mathbf{g}_s \in \mathbb{C}^{1 \times N_r}$ is the receiver filter; $\mathbf{H}_{i,s} \in \mathbb{C}^{N_r \times N_t}$ represents the small scale fading; $\mathbf{f}_i \in \mathbb{C}^{N_t \times 1}$ is the transmit precoder; P_i is the transmit power; and σ_0^2 is the total background thermal noise power.

B. BASELINE CQI MEASUREMENT
In the baseline approach, the UE's CQI report on the n^*-th TTI (with n^* satisfying $n^* \bmod N = 0$) is based on the latest available SINR measurement $\Psi_s[n]$ at the moment of the report, where $n^* > n$ due to the processing time needed to prepare the CQI report. The CQI processing times at the UE, its transmission to the gNB, and the corresponding gNB CQI decoding, means that the CQI applied for a link adaptation (and scheduling) decision on TTI n' is representing the quality at the UE in TTI $n' - w$, where w corresponds to the effective CQI processing and reporting delay offset. This means that the channel quality assumed by the gNB for its link adaptation (and scheduling) decision may not match the channel quality of the actual transmission. This problem is exacerbated in the considered scenario with short and frequent TSC transmissions at each cell, in which the experienced interference at the UE is subject to large variations from one TTI to another TTI. Such interference variations are mainly originating from changes in RB allocations of the neighboring cells, and use of different transmitter precoders, depending on which UEs (and their channel conditions) are scheduled in the neighboring cells.

C. LOW-PASS FILTERED CQI MEASUREMENT

To address the problem of the rapidly-changing interference pattern, it is proposed to have low-pass filtering of the interfering measurements at the UE (i.e. the denominator of (4)) that are used for the SINR estimation [13]. On each TTI n, the UE measures the interference I_s calculated as

$$I_s[n] = \sum_{i \in \mathcal{I}} \Omega_i[n] \|\mathbf{g}_s[n]\mathbf{H}_{i,s}[n]\mathbf{f}_i[n]\|^2 P_i[n] + \sigma_0^2. \quad (5)$$

The $I_s[n]$ measurements are filtered with a low-pass first-order infinite-impulse response (IIR) filter, resulting in the following smoothed interference value:

$$I_s'[n] = \alpha \cdot I_s[n] + (1 - \alpha) \cdot I_s'[n-1] , \quad (6)$$

where α is the forgetting factor of the filter ($0 < \alpha < 1$). On each CQI reporting opportunity at TTI n^*, the CQI is determined according to (3), with Ψ_s consisting of the ratio between the latest desired-signal fading information and the low-pass filtered interference $I_s'[n]$:

$$\Psi_s[n] = \frac{\Omega_c[n] \|\mathbf{g}_s[n]\mathbf{H}_{c,s}[n]\mathbf{f}_c[n]\|^2 P_c}{I_s'[n]} . \quad (7)$$

The value of α determines how much weight is given to the latest interference measurement as compared to the previous ones. Following the previous work in [13], we assume $\alpha = 0.01$.

D. gNB-BASED CQI ADJUSTMENT

As a second enhancement for DPS, the received CQI reports at the gNB are offset by a safety factor $\Delta > 0$ [dB] to account for the potential mismatch between the received CQI and the channel quality at the UE. This is needed in order to ensure the required reliability of the packet transmission, since the stringent latency requirements of TSC does not leave room for recovery through HARQ retransmissions, as mentioned previously. Note that large values of Δ provide improved reliability, however it leads to reduced spectral efficiency, which may limit the number of users that can be served in the network. The offset Δ is subtracted from Ψ_s (in SINR domain) before selecting the MCS. This is feasible since the CQI table is designed to have constant SINR offset between the entries; thus, if the SINR offset between the CQI table entries is ϕ dB, the received CQI index is offset by $\lfloor \Delta/\phi \rceil$ before being used for determining the MCS.

IV. ENHANCEMENTS FOR SEMI-PERSISTENT SCHEDULING

For SPS, the gNB sets the long-term MCS and RB allocation for the periodical DL transmissions that ensures the required reliability for future transmissions to the UE. This essentially means that link adaptation for SPS will be slow, as the gNB must select the MCS for the SPS allocations in coherence with the experienced SINR statistics at the UE when observed over a large number of periodical resource allocations. Notice that the use of slow link adaptation for SPS also implies that the potential benefits from fast radio channel-aware link

adaptation and scheduling (as known from DPS with fast link adaptation) are sacrificed for the benefit of not having control channel overhead (DCI) for every DL transmission.
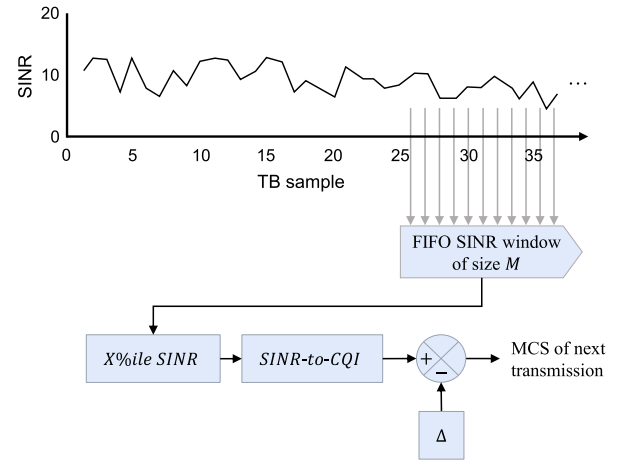

FIGURE 2. SINR windowing for MCS selection.

A. MCS ADJUSTMENT

The following two approaches for MCS selection are studied:

1) In the first approach, a fixed MCS is defined for all the UEs in the network (as also adopted in [16]). This is the simplest approach, as the gNB configures all the UEs with the same MCS regardless of their individual channel conditions. However, it tends to be spectral inefficient as the MCS must be chosen such that the high reliability requirements can be guaranteed to even the UEs with the worst SINR conditions.

2) The second approach is based on a *semi-static* MCS selection scheme. As illustrated in Fig. 2, the MCS is determined based on the SINR samples of the previously transmitted transport block (TB) transmissions. For each received TB, the UE estimates the SINR over the allocated RBs and stores the value in a first-in-first-out (FIFO) buffer of size M. A CQI is determined according to (3), with Ψ_s corresponding to the X-th percentile of the stored SINR observations. A single CQI ($s = 1$; $y = Y$) is reported if the corresponding MCS index changes. Similarly as for DPS, the gNB may subtract an offset $\Delta > 0$ [dB] from the CQI to account for potentially worse SINR conditions than what has been experienced in the past observation interval.

Notice that for the second method the MCS is occasionally updated in case the reported CQI, discounting the offset given by Δ, indicates a different index. Thus it allows each UE configuration to be updated according to its experienced channel condition. However, if the MCS is frequently updated, the impact of the control signaling needed for the update can get as high as with the DPS approach. Therefore, the windowing parameter M should be selected in order to

minimize the need for reconfiguring the SPS parameters. A small M value leads to less historic SINR information and potentially frequent MCS updates for UEs with high channel variability, e.g. cell-edge UEs more susceptible to inter-cell interference. On the other hand, a large M value allows a longer SINR backlog with the cost of more memory usage and slower adaptation to the changing channel conditions. The percentile value X can be configured, for example, to track median SINR performance with $X = 50\%$ile, or a very low X can be used to track worst-case SINR values. The impact of these parameters will be examined in Section VI.

B. RB ALLOCATION

Given the selected MCS for an SPS configuration, the number of RBs is selected to deliver a TB containing one TSC packet on every configured resource allocation opportunity. As mentioned, frequency-domain channel-aware RB allocation is not supported for SPS. Three RB allocation options for SPS are considered, namely *sequential RBs*, *shuffled RBs* and *soft RBs split*.

1) With a *sequential* RBs allocation, each gNB starts allocating from the first available RB to the last available RB in the bandwidth part in a sequential fashion. This is the simplest yet naive approach, as it can lead to high inter-cell interference since all the cells will start the allocation from the same RB.

2) In the *shuffled* RB approach, the RBs for the UEs in each cell are allocated in a random order, thus reducing the chance of co-channel interference among the neighbor cells.

3) In the *soft* RBs split approach, a simple coordination scheme ensures that each cell starts allocating the RBs from a certain point in the frequency domain. The starting RB index j of a cell, with maximum $Y - 1$, can be determined based on the gNB index $c \in \mathcal{C}$ assuming sequential indexing and a split factor $k \in \{1, \ldots, C\}$ as follows:

$$j = \lfloor c \cdot Y / k \rfloor \mod Y. \tag{8}$$

Figure 1 shows an example for $k = 4$, where cells with the same allocation starting point have the same area color.

For all options, a resource allocation equivalent to frequency reuse factor of 1 is assumed. This means that each cell may allocate up to Y RBs if required, and can potentially cause interference on neighbor cells.

V. EVALUATION METHODOLOGY

We evaluate the performance of a TSC deployment through extensive dynamic system-level simulations. The simulator includes explicit modeling of the RAN user plane protocols, radio resource management (RRM) mechanisms, 3D radio propagation, and traffic models commonly accepted for evaluation in 3GPP. Table 2 summarizes the main simulation

TABLE 2. Summary of simulation assumptions.

Parameters	Assumption
Layout	Indoor office, 12 gNBs per 120 m x 50 m
Inter-site distance	20 m
Carrier frequency	4 GHz
Channel model	ITU InH for 4 GHz
UE distribution	Uniformly distributed indoor, speed 3 km/h
gNB transmitter	$N_t = 4$ antennas in 2 columns and two polarizations, 24 dBm
gNB antenna gain	5 dBi
gNB antenna height	10 m (wide beam pointing downwards)
UE receiver	$N_r = 4$ antennas, MMSE-IRC receiver
UE noise figure	9 dB
Thermal noise	-174 dBm/Hz
Bandwidth	20 MHz for the DL, FDD
PHY numerology	30 kHz SCS, 12 sub-carriers/RB, 2-symbols mini-slot
Traffic model	20 bytes and 50 bytes payload, 0.5 ms periodicity, random arrival time offset per flow
HARQ/repetition	Retransmissions and repetitions disabled
Overhead	PDCCH overhead explicitly modelled
DPS baseline CQI	Frequency-selective CQI with sub-band size $y = 4$ RBs and reporting every 2 ms ($N = 28$ TTIs)
DPS CQI with filtering	Same as baseline, but with IIR filtered interference for UE CQI estimation with filtering coefficient $\alpha = 0.01$.
SPS CQI	Wideband CQI according to the X-th percentile of SINR samples from up to $M = 500$ past TB allocations ($X = 0.1$ for minimum and $X = 50$ for median). CQI is reported when the corresponding MCS index changes.

assumptions. The assumptions are based on NR evaluation methodology for Release-16 considering the factory automation scenario as summarized in [29].

An indoor deployment as illustrated in Fig. 1 is assumed, comprising 12 gNBs with antennas located at 10 m height, and panels pointing downwards. These gNBs are intended to provide the required coverage for the 50 m x 10 m service areas as described in Table 1 within the 120 m x 50 m floor area. The carrier frequency is 4 GHz and a dedicated bandwidth of 20 MHz containing $Y = 50$ RBs is assumed for the DL transmissions using FDD. The UEs are uniformly distributed in the indoor environment with semi-stationary positions, i.e. mobility/handover is not considered while a 3 m/s UE speed is assumed for modeling fast fading effects.

Each gNB has a panel with $N_t = 4$ transmit antennas and each UE utilizes $N_r = 4$ receive antennas, creating a 4×4 closed loop single-user multiple input multiple output (SU-MIMO) system. Rank one transmissions are assumed. A minimum mean square error - interference rejection combining (MMSE-IRC) receiver is assumed at the UE side following the model presented in [30]. The receiver filter **g** for a desired signal in the UE is given by

$$\mathbf{g} = \mathbf{f}_c^H \mathbf{H}_c^H \mathbf{R}^{-1}, \tag{9}$$

where $(\cdot)^H$ is the Hermitian operator and $\mathbf{R}$ is the IRC interference covariance matrix,

$$\mathbf{R} = P_c \mathbf{H}_c \mathbf{f}_c \mathbf{f}_c^H \mathbf{H}_c^H + \sum_{i \in \mathcal{I}} P_i \mathbf{H}_i \mathbf{f}_i \mathbf{f}_i^H \mathbf{H}_i^H + \sigma_0^2 \mathbf{I}_{N_r}, \quad (10)$$

where $\mathbf{I}_{N_r}$ is the identity matrix. The mutual-information-based effective SINR mapping is used for determining whether a transport block is received [31]. The post-receiver SINR is calculated as in (4) for each RE of the RBs allocated for the transmitted packet (for this reason, the sub-band index s is omitted here). The SINR values are combined in the mutual information domain to form an effective SINR value. The effective SINR is then mapped to a correspondent BLEP value according to the applied MCS.

The periodicity of the TSC traffic flows for each UE is $T = 0.5$ ms. Two payload sizes are considered, $B = 20$ bytes and $B = 50$ bytes. Each periodic traffic flow has a random arrival time offset. The target is transmitting the TSC packets within $D = T = 0.5$ ms with 99.999% success probability – note that the latency deadline is equivalent to the traffic periodicity T as shown in Table 1. The traffic is scheduled according the mechanisms described in Section III and Section IV.

We adopt the definition of reliability described in [32], which is stated as the percentage value of the amount of network layer packets successfully delivered within the time constraint, as a fraction of the total number of sent network layer packets. In order to determine the network capacity supported with each scheduling approach and TSC traffic, a series of simulations varying the load with an increasing number of UEs are conducted. The network capacity is found by the maximum supported number of TSC UEs where the TSC latency and reliability requirements can be satisfied. For ensuring statistical reliable results, the simulations are run to collect at least 5 million packet transmissions in the network. Each simulation consists of 10 drops in which the UEs are redistributed in random positions. Therefore, the experienced SINR for each UE transmission varies due to many factors, including path-loss and shadow fading which changes in each drop, fast fading effects, the offered traffic load, as well as the RB allocation and precoding of the transmissions at the interfering cells.

The control channel resources for DPS are dynamically adjusted for meeting a reliability target ten times higher than the data, i.e. the control channel error probability is in the order of 10^{-6}. The overhead expressed in terms of number of CCEs is determined according to the reported wideband CQI value ($y = Y$), as shown in Table 3. We note that the SINR level is typically high in the considered indoor scenario (in comparison to e.g. macro deployments), incurring that a single CCE is typically sufficient to meet the target reliability.

VI. PERFORMANCE RESULTS

The performance of the TSC network is evaluated for the different scheduling strategies and packet sizes B. The network capacity is compared in terms of the number of supported TSC UEs.

TABLE 3. Control channel overhead depending on channel quality.

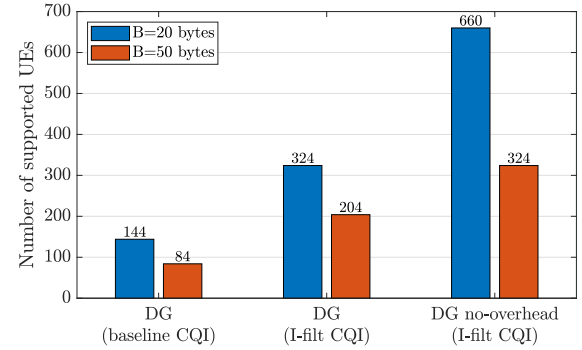| Wideband CQI [dB]	#CCE	Control channel overhead
$[4.2, \infty)$	1	$1 \times 72 = 72$ REs
$[0.2, 4.2)$	2	$2 \times 72 = 144$ REs
$[-2.2, 0.2)$	4	$4 \times 72 = 288$ REs
$(-\infty, -2.2)$	8	$8 \times 72 = 576$ REs

FIGURE 3. Network capacity in terms of number of supported TSC UEs with DPS.

A. DPS-ENABLED TSC

First we analyse the number of supported UEs when using DPS, considering the impact of the control channel overhead which is present for every data transmission. The value of safety margin offset Δ is determined empirically for each case by varying it by 1 dB from 0 to 6 dB and selecting the value which leads to the highest network capacity. Figure 3 shows the number of supported UEs with DPS using baseline CQI and using interference filtered CQI (labelled as I-Filt CQI). In addition, we show also a reference case where the control channel impact is neglected (i.e., an artificial case where the control overhead is zero). The frequent small packet transmissions in the considered dense indoor deployment leads to high interference variability. Thus, DPS based on the baseline CQI achieves a limited TSC network capacity. On the other hand, DPS based on interference filtered CQI allows to more than double the TSC network capacity compared with using baseline CQI. This is due to the more accurate link adaptation, as also discussed in [13]. As expected, the number of supported UEs is lower when B increases from 20 to 50 bytes; however, despite the 2.5x increase of B, the relative decrease of the number of UEs is much smaller (between 1.6 and 1.7 for Baseline and I-filt CQI, respectively) since the TSC network capacity is to a large extent constrained by the control channel overhead. This is reflected when looking at the case where the overhead is artificially removed, where network capacity is more than doubled for the $B = 20$ bytes case and approximately 60% larger for $B = 50$ bytes case.

Figure 4 shows further evidence on the impact of the control channel overhead. Specifically, the cumulative distribution function (CDF) of the allocation size (in number of RBs) and the relative control channel overhead is depicted for the case of interference filtered CQI. The number of allocated RBs is naturally higher for transmissions with a larger payload size, whereas the relative overhead decreases
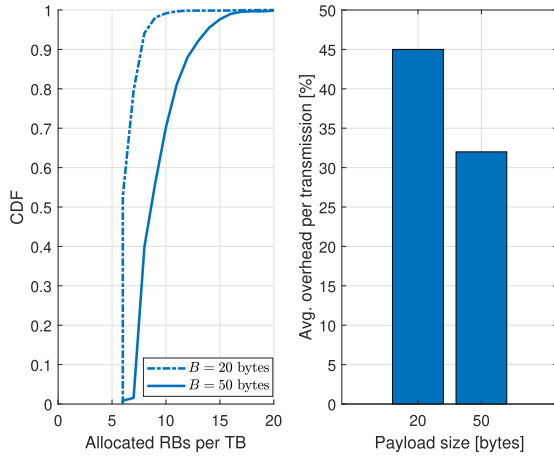
FIGURE 4. DPS allocation size and overhead per transmission.

FIGURE 5. SINR for different RB allocation strategies.

FIGURE 6. Network capacity in terms of number of supported TSC UEs with SPS.

as the number of resources on the PDCCH does not depend on the scheduled TB size. We remark that the control channel not only occupies resources, but it also loads the network causing additional interference thus impacting considerably the achievable performance.

B. SPS-ENABLED TSC

We start by evaluating the three RB allocation strategies for SPS described in Sec. IV-B. Fig. 5 shows the CDF of the SINR [1] of the transmitted TBs for the different RB allocation approaches and for low and high offered load conditions of TSC traffic. *High load* corresponds to the highest load that the system can handle while still achieving the TSC requirements; that is, 660 UEs with $B = 20$ byte payload size corresponding to 32% RB utilization.

It is clear that the sequential approach is inadequate given the poor SINR performance for both offered loads. This is due to the high interference from neighbor cells which are constantly allocating the same resources in frequency domain, as in a fully loaded system.

On the other hand, the shuffled RBs approach improves the SINR performance up to 20 dB. That is possible due to the low RB utilization typical of TSC and URLLC systems, which operate with fractional load. Finally, the soft RBs split approach with $k = 4$ provides 1 to 2 dB higher SINR compared to shuffled RB allocation, for low and high loads. This further improvement can be achieved because the gNBs utilizing the same allocation starting point in the scenario are far apart as illustrated in Fig. 1 (each color represents the gNB utilizing the same allocation starting point). In light of the observed performance, in the next results we assume the soft RBs split for SPS.

Now we evaluate the achieved network capacity obtained by SPS with the MCS adjustments proposed in Sec. IV-A; the simulation results are shown in Fig. 6 for 20 bytes and
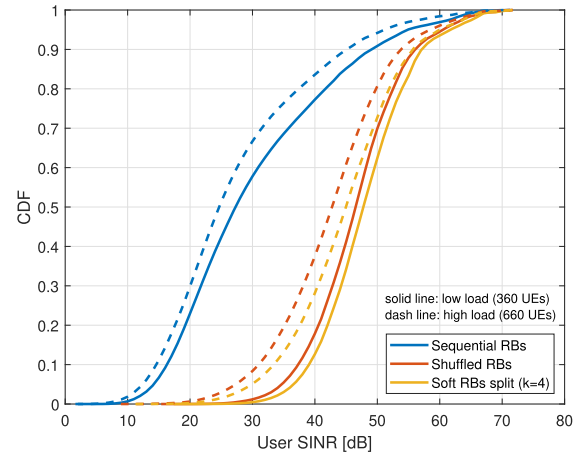
50 bytes payload. For the case where a fixed MCS is assigned to all UEs, the MCS was determined empirically as the MCS which leads to the highest network capacity with each packet size. QPSK 2/3 is found to be the best option for 20 bytes payload size and 16QAM 1/2 for 50 bytes payload size. Besides, cases with semi-static MCS adjustment based on the median SINR and minimum SINR from a sample window of size $M = 500$ are shown as well.

The achieved network capacity of SPS with fixed MCS is lower compared with SPS operation with semi-static MCS selection. However, it is relevant to observe that it is still superior to DPS based on baseline CQI, mainly due to the lack of control channel overhead. Moreover, it can be seen that the semi-static MCS selection based on median SINR from the collected samples does not track well the worst case SINR of each users, thus requiring a high safety margin offset (in this case, $\Delta = 6$ dB). The highest network capacity is achieved with a MCS selection based on the minimum SINR, which is able to track the worst case condition and a less conservative offset of $\Delta = 1$ dB is applied.

As mentioned in Section III, the SPS allocation parameters can be updated by a DCI. However, frequent updates would lead to the same issues as present in DPS such as control signaling overhead and unreliability. In Fig. 7 we observe the time between updates for the semi-static MCS schemes and the case of DPS without control overhead as a comparative

---

[1]The minimum user SINR is relatively high (approx. 10 dB) because here we consider the *post-receiver* SINR, which accounts for precoding and interference-rejection combining (IRC) gains given by the multi-antenna configuration.
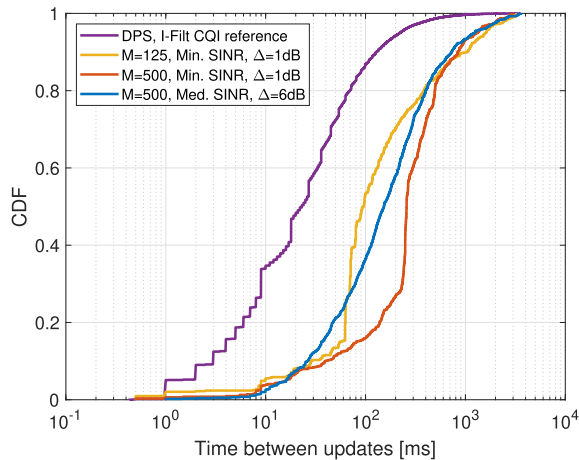
FIGURE 7. Time between MCS updates for SPS with different SINR windowing settings and for reference DPS scheme (660 UEs, 20 bytes).

reference. It can be noticed that the semi-static MCS approach updates 10 to 15 times less frequent compared with the DPS baseline. Moreover, using minimum SINR is not only superior than median SINR in terms of achieved network capacity, but also has the lowest MCS update frequency. In 85% of the cases, the MCS is updated after more than 100 ms, which translates to a very low control signaling overhead and therefore better performance.

VII. CONCLUSION

5G-enabled Time-Sensitive Communications (TSC) enables more efficient wireless support for deterministic periodic traffic flows thanks to the introduction of TSC Assistance Information (TSCAI), which can be leveraged by the packet scheduler to tailor resource allocation to the application. In this paper, we have developed and assessed both dynamic packet scheduling (DPS) and semi-persistent scheduling (SPS) strategies. For DPS-enabled TSC, we utilize the UE based measuring and reporting scheme from [13] and apply an additional safety offset parameter to account for the critical SINR variations that may otherwise jeopardize TSC performance. This method provides more than 3x gain over using traditional DPS methods proposed for URLLC traffic. For SPS-enabled TSC, a new UE based measuring and reporting scheme is proposed, where measurements are conducted on actual data allocations over a time window. The optimized method uses the minimum observed SINR with an offset that is applied at the gNB for MCS selection. Moreover, a proposed soft frequency-reuse allocation method between cells provides approximately 2 dB SINR gain over a random allocation of users in each cell. Compared to a configured fixed MCS reference scenario, the new scheme provides 3x gains for 20 bytes payload and 2x gains for a 50 bytes payload. While DPS has the benefit of adapting the allocation to the current channel conditions and leveraging frequency selective scheduling, the absence of HARQ and the control channel overhead limits its performance. A better performance is found using SPS with

the proposed semi-static MCS selection scheme leveraging the TSCAI where the signalling overhead is shown to be manageable. The proposed SPS scheme supports 2x more users for 20 bytes payload and 76% more users for a 50 bytes compared to DPS.

REFERENCES

[1] C. Stefanovic, "Industry 4.0 from 5G perspective: Use-cases, requirements, challenges and approaches," in *Proc. 11th CMI Int. Conf.*, Copenhagen, Denmark, Nov. 2018, pp. 44–48.

[2] S. Vitturi, C. Zunino, and T. Sauter, "Industrial communication systems and their future challenges: Next-generation Ethernet, IIoT, and 5G," *Proc. IEEE*, vol. 107, no. 6, pp. 944–961, Jun. 2019.

[3] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. ElBakoury, "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 88–145, 1st Quart., 2019.

[4] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[5] V. K. L. Huang, Z. Pang, C.-J.-A. Chen, and K. F. Tsang, "New trends in the practical deployment of industrial wireless: From noncritical to critical use cases," *IEEE Ind. Electron. Mag.*, vol. 12, no. 2, pp. 50–58, Jun. 2018.

[6] *NR; NR and NG-RAN Overall Description*, document 38.300 v16.0.0, 3GPP, Dec. 2019.

[7] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, and Z. Li, "Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements," *IEEE Netw.*, vol. 32, no. 2, pp. 8–15, Mar. 2018.

[8] *Service Requirements for Cyber-Physical Control Applications in Vertical Domains*, 3GPP, document 22.104 v17.0.0, Jun. 2019.

[9] *Study on NR Industrial Internet of Things (IoT)*, document 38.825 v6.0.0, 3GPP, Mar. 2019.

[10] *System Architecture for the 5G System (5GS)*, document 23.501 v16.3.0, 3GPP, Dec. 2019.

[11] A. Alabbasi, T. Dudda, Z. Zou, and J. Kronander, "5G toolbox for realizing industrial automation," in *Proc. IEEE 2nd 5G World Forum (5GWF)*, Sep. 2019, pp. 512–515.

[12] *Summary of RAN1 NR IIoT WI agreements & conclusions up to RAN1#99*, document R1-1913602, 3GPP, Nov. 2019.

[13] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC layer enhancements for ultra-reliable low-latency communications in cellular networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2017, pp. 1005–1010.

[14] A. Belogaev, E. Khorov, A. Krasilov, D. Shmelkin, and S. Tang, "Conservative link adaptation for ultra reliable low latency communications," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, Jun. 2019, pp. 1–5.

[15] D.-T. Phan-Huy, P. Chauveau, A. Galindo-Serrano, and M. Deghel, "High data rate ultra reliable and low latency communications in bursty interference," in *Proc. 25th Int. Conf. Telecommun. (ICT)*, Jun. 2018, pp. 175–180.

[16] M. Condoluci, T. Mahmoodi, E. Steinbach, and M. Dohler, "Soft resource reservation for low-delayed teleoperation over mobile networks," *IEEE Access*, vol. 5, p. 10445–10455, May 2017.

[17] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 127–141, Jan. 2018.

[18] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2007, pp. 2861–2864.

[19] M. Rinne, M. Kuusela, E. Tuomaala, P. Kinnunen, I. Kovacs, K. Pajukoski, and J. Ojala, "A performance summary of the evolved 3G (E-UTRA) for voice over Internet and best effort traffic," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3661–3673, Sep. 2009.

[20] G. Karadag, R. Gul, Y. Sadi, and S. Coleri Ergen, "QoS-constrained semi-persistent scheduling of machine-type communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2737–2750, May 2019.

[21] A. G. Gotsis, A. S. Lioumpas, and A. Alexiou, "Evolution of packet scheduling for machine-type communications over LTE: Algorithmic design and performance analysis," in *Proc. IEEE Globecom Workshops*, Dec. 2012, pp. 1620–1625.

[22] S.-Y. Lien and K.-C. Chen, "Massive access management for QoS guarantees in 3GPP Machine-to-Machine communications," *IEEE Commun. Lett.*, vol. 15, no. 3, pp. 311–313, Mar. 2011.

[23] *Radio Resource Control (RRC) Protocol Specification*, document 38.331 v15.9.0, 3GPP, Mar. 2020.

[24] *Physical Layer Procedures for Control*, document 38.213 v16.0.0, 3GPP, Dec. 2019.

[25] N. Nokia and S. Bell, *Discussion on the RAN2 LS on TSN Requirements Evaluation*, document R1-1813120, 3GPP, Nov. 2018.

[26] *Physical Layer Procedures For Data*, document 38.214 v16.0.0, 3GPP, Dec. 2019.

[27] R. Agarwal, V. Majjigi, Z. Han, R. Vannithamby, and J. Cioffi, "Low complexity resource allocation with opportunistic feedback over downlink OFDMA networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1462–1472, Oct. 2008.

[28] T.-S. Kang and H.-M. Kim, "Opportunistic feedback assisted scheduling and resource allocation in OFDMA systems," in *Proc. 10th IEEE Singap. Int. Conf. Commun. Syst.*, Mar. 2006, pp. 1–5.

[29] *Study on Physical Layer Enhancements for NR Ultra-Reliable Cand Low Latency Case (URLLC)*, document 38.824 v16.0.0, 3GPP, Mar. 2019.

[30] K. Pietikainen, F. Del Carpio, H.-L. Maattanen, M. Lampinen, T. Koivisto, and M. Enescu, "System-level performance of interference suppression receivers in LTE system," in *Proc. IEEE 75th Veh. Technol. Conf. (VTC Spring)*, May 2012, pp. 1–7.

[31] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, *IEEE 802.16m Evaluation Methodology Document (EMD)*, document 802.16m-08/004r2, Jul. 2008.

[32] *Service Requirements for the 5G System*, 3GPP, document 22.261 v16.9.0, Sep. 2019.

THOMAS H. JACOBSEN (Member, IEEE) received the M.Sc. and Ph.D. degrees from Aalborg University, Denmark, in 2015 and 2019, respectively.

He is currently with Nokia Bell Labs as a Device Standardization Research Expert, where his main research focus on uplink control information and radio resource management for ultrareliable and low-latency communications, as well as time synchronization for time sensitive communications and the Industrial Internet of Things use cases with 5G new radio.

MARCO CENTENARO (Member, IEEE) received the B.S. degree in information engineering and the M.S. and Ph.D. degrees in telecommunication engineering from the University of Padova, Italy, in 2012, 2014, and 2018, respectively.

He was on leave as a Visiting Researcher with Nokia Bell Labs, Stuttgart, Germany, and Yokohama National University, Japan, in 2016 and 2017, respectively. From 2018 to 2019, he was Postdoctoral Research Fellow with the University of Padova, where he is currently doing research on projects funded by Huawei Technologies, and with Aalborg University, Denmark, where he also working in tight collaboration with Nokia Bell Labs Aalborg. At the time of writing, he is also an Expert Research with Fondazione Bruno Kessler, Italy. His researches are focused on telecommunication standards and the Internet of Things technologies.

RENATO B. ABREU (Member, IEEE) received the degree in electrical engineering from the Federal University of Viçosa, Brazil, in 2007, the M.Sc. degree from the Federal University of Amazonas, in 2014, and the Ph.D. degree in wireless communication networks from Aalborg University, Denmark, in 2019.

From 2008 to 2016, he worked with Test Engineering and Connectivity Research at the former Nokia Development Institute (INDT), Brazil. Since 2019, he has been working as a Device Standardization Expert with Nokia Bell Labs. His research interests include ultrareliable and low-latency communications, and the Industrial Internet of Things in 5G networks.

KLAUS I. PEDERSEN (Senior Member, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively. He is currently leading the Nokia Bell Labs radio access systems research team in Aalborg, and a part-time Professor with the Wireless Communications Network (WCN) Section, Aalborg University. He is the author or coauthor of approximately 200 peer-reviewed publications on a wide range of topics, as well as inventor on several patents. His current work is related to 5G new radio evolution, including radio resource management aspects to enable new use cases with special emphasis on mechanisms that offer improved end-to-end (E2E) performance delivery. He was recently also part of the EU funded research project ONE5G that focused on E2E-aware Optimizations and advancements for the Network Edge of 5G New Radio that was successfully concluded in the Summer 2019.

GUILLERMO POCOVI (Member, IEEE) received the M.Sc. degree in telecommunications engineering from the Universitat Politècnica de Catalunya, in 2014, and the Ph.D. degree from Aalborg University, Denmark, in 2017. He is currently an Industrial Postdoctoral Researcher with Nokia Bell Labs Aalborg, partly sponsored by the Innovation Fund Denmark (IFD) under File No. 7039-00009B. His research activities are related to the support of ultrareliable and low-latency communications and the Industrial Internet of Things (IIoT) use cases in 5G new radio.

TROELS E. KOLDING (Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunications from Aalborg University, Denmark, in 1996 and 2000, respectively. He has been working in various research and management positions at Nokia, since 2001. He is currently a Senior Specialist with Nokia Bell Labs. His current research responsibilities include 5G time-sensitive communications, wireless time sensitive networking, cloud-based network architectures, and time synchronization.

● ● ●